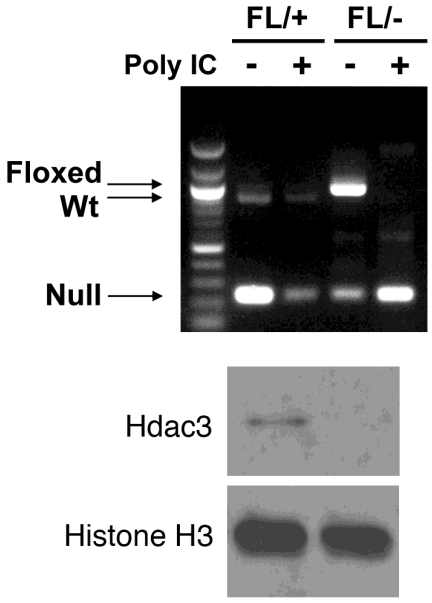
SI Appendix

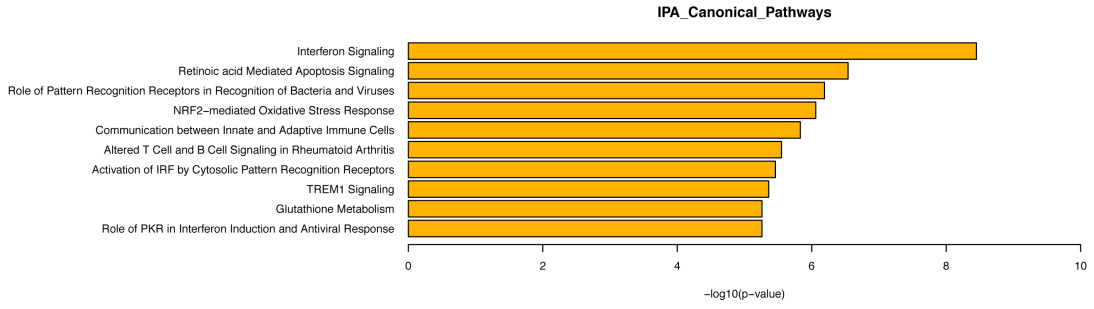
Requirement for the histone deacetylase Hdac3 for the inflammatory gene expression program in macrophages (Chen et. al. 2012)

Figures



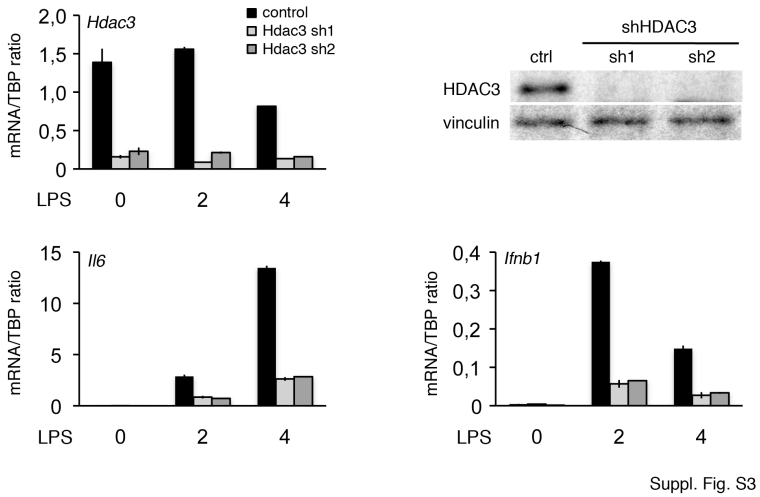
Suppl. Fig. S1

Suppl. Fig. S1. *Deletion efficiency in Mx-Cre/Hdac3fl/fl mice*. Top: Agarose gel showing the genotyping procedure. Bottom: western blot analysis. Histone H3 is used as a loading control.



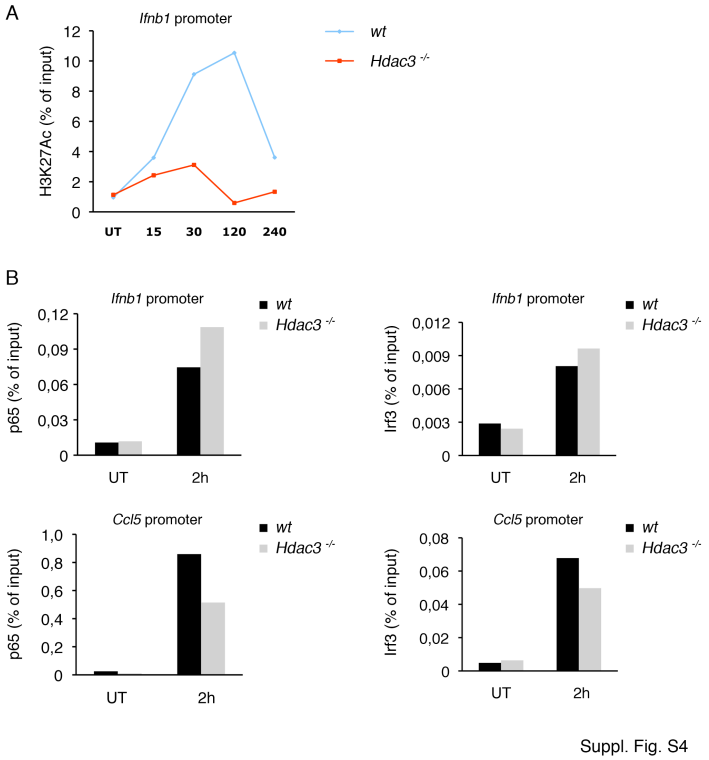
Suppl. Fig. S2

Suppl. Fig. S2. *Ingenuity pathway analysis in Hdac3-null macrophages stimulated with LPS*.



Suppl. Fig. S3

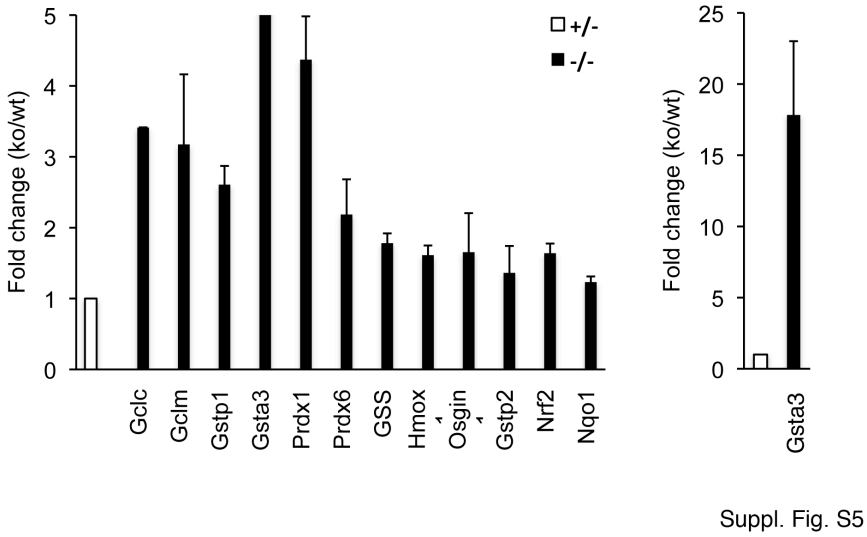
Suppl. Fig. S3. *Impaired inflammatory gene expression in HDAC3-depleted macrophages*. Differentiating bone marrow cells were infected with retroviruses harboring two different shRNAs targeting Hdac3.



Suppl. Fig. S4

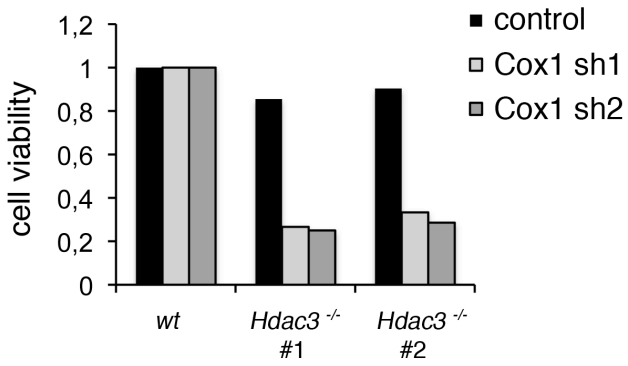
Suppl. Fig. S4. *p65/RelA and Irf3 recruitment to the Ifnb1 promoter in Hdac3-/- macrophages*. **a**) Histone H3K27 acetylation at the *Ifnb1* promoter in wt and Hdac3-

deficient cells. **b**) p65 and Irf3 recruitment to the Ifnb1 promoter in Hdac3-deficientmacrophages. The Ccl5 promoter is shown as control.



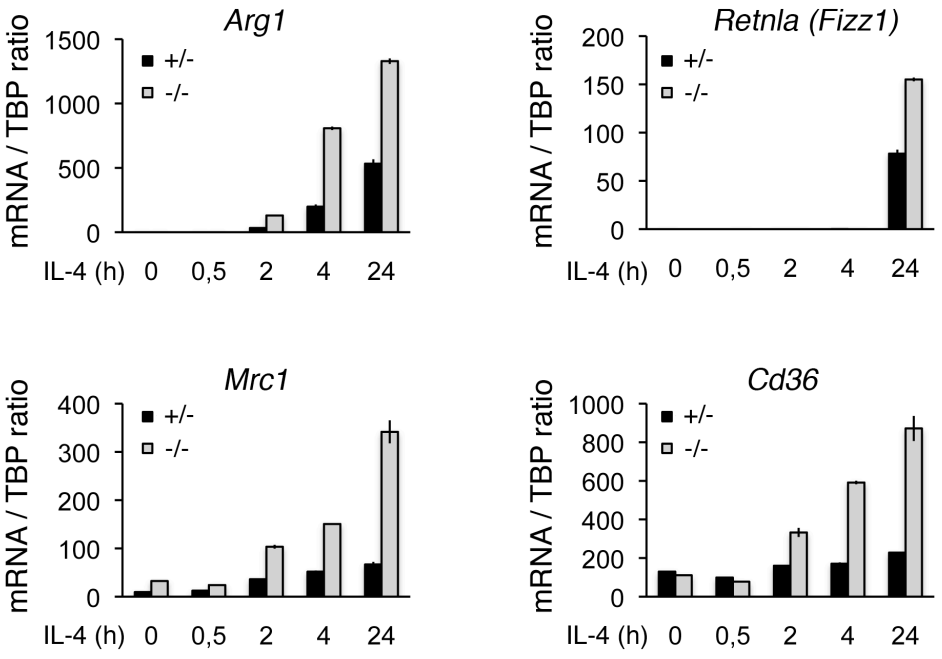
Suppl. Fig. S5

Suppl. Fig. S5. *Nrf2-target genes upregulated in Hdac3$^{-/-}$ macrophages*. Changes in expression of the indicated Nrf2-dependent genes was measured by quantitative RT-PCR.



Suppl. Fig. S6

Suppl. Fig. S6. *Depletion of Cox-1 impairs viability of Hdac3-/- cells.* Data obtained with two independent preparations of *Hdac3 $^{-/-}$* cells are shown.



Suppl. Fig. S7

Suppl. Fig. S7. *Expression of M2 genes in Hdac3-deficient macrophages*.

Datasets and Tables

Dataset S1. *cDNA microarray data in untreated or LPS-treated Hdac3-/- macrophages.*

Dataset S2. *Ingenuity pathway analysis*.

Dataset S3. *PSCAN analysis for transcription factor binding site overrepresentation in differentially expressed genes*.

Dataset S4. *Acetyl-H4 ChIP-Seq data*.

Table S1. Clover analysis hyper- and hypo-acetylated regions detected by ChIP-Seq.

Hyperacetylated Distal LPS

PWM	score	p
MA0099.2_AP1	570	0
MA0099.1_Fos	387	0
BU0047_Myf6_primary	332	0.006
MA0095.1_YY1	189	0
MA0112.1_ESR1	156	0.009
MA0071.1_RORA_1	124	0
JMB_CAATT_box	39.7	0
TA0012_RFX3_dimer	28.7	0.01
BU0098_Zfp410_primary	26.1	0.002
MA0078.1_Sox17	7.04	0
MA0066.1_PPARG	6.99	0
MA0117.1_Mafb	6.88	0.005
TA0020_TFEB_dimer	0.545	0.001
MA0072.1_RORA_2	-0.85	0.002
MA0058.1_MAX	-1.35	0
MA0089.1_NFE2L1_MafG	-3.73	0
MA0147.1_Myc	-3.77	0
MA0104.2_Mycn	-4.03	0
MA0067.1_Pax2	-4.58	0.003
MA0093.1_USF1	-4.76	0
MA0104.1_Mycn	-4.77	0
TA0022_XBP1_monomer	-4.94	0
BMC_DPE	-5.55	0.002
MA0035.1_Gata1	-5.8	0.003
MA0004.1_Arnt	-5.85	0
MA0037.1_GATA3	-6.4	0.008
MA0036.1_GATA2	-6.62	0.001
MA0259.1_HIF1A_ARNT	-7.79	0.004

Hyperacetylated distal UT

PWM	score	p
MA0442.1_SOX10	213	0
MA0145.1_Tcfcp2l1_	92.8	0.002
BU0104_Zscan4_primary	46.5	0.004
TA0007_GLI2_monomer	45.7	0.006
MA0095.1_YY1	19.5	0
TA0029_GLI3	14.6	0.004
MA0117.1_Mafb	6.5	0.006
HOMEO0113_Nkx2-2_2823.1	4.1	0.003
BU0041_Mafb_primary	3.68	0.004

Hypoacetylated distal LPS

PWM	score	p
MA0137.1_STAT1	444	0
MA0050.1_IRF1	289	0
MA0051.1_IRF2	88.7	0
BU0037_Isgf3g_primary	80.6	0.004
MA0112.1_ESR1	63.6	0.006
MA0095.1_YY1	53.3	0
MA0143.1_Sox2	35.2	0.003
HOMEO0084_Irx2_0900.3	2.83	0.001
HOMEO0088_Irx5_2385.1	2.56	0.005
HOMEO0089_Irx6_2623.2	2.24	0.01
MA0078.1_Sox17	-0.0467	0.01
HOMEO0028_Duxl_1286.2	-4.97	0.01

Hypoacetylated distal UT

PWM	score	p
ETS0004_h-ELF3	211	0.004
ETS0001_h-EHF	121	0.005
ETS0020_h-ETV7	111	0.008
MA0062.1_GABPA	98.8	0.01
MA0139.1_CTCF	55	0
MA0122.1_Nkx3-2	28.7	0.001
ETS0015_h-ETV2	20.8	0.006
MA0062.2_GABPA	20.2	0.003
BU0101_Zic1_primary	4.66	0.002
BU0102_Zic2_primary	4.23	0
BU0103_Zic3_primary	4.15	0
BU0038_Jundm2_primary	-1.65	0.004
BU0004_Atf1_primary	-2.23	0.005
ETS0005_h-ELF4	-2.26	0.01
MA0009.1_T	-3.08	0.005

Table S2. *Primers used in the study*.

Gene symbol	Refseq	Primers
ActB	NM_007393	CCCTGAAGTACCCCATTGAA
		GGGGTGTTGAAGGTCTCAAA
Ccl5	NM_013653	ACCATATGGCTCGGACACCACT
		ACCCACTTCTTCTCTGGGTTGG
Ifnb1	NM_010510	GCTCCAAGAAAGGACGAACA
		CCCAGTGCTGGAGAAATTGT
Il6	NM_031168	CCATAGCTACCTGGAGTACATG
		TGGAAATTGGGGTAGGAAGGAC
Nos2	NM_010927	CCATCATGAACCCCAAGAGT

		CATCCAGAGTGAGCTGGTAGG
Ptgs1	NM_008969	TGCCCTCTGTACCCAAAGAC
		TGTGCAAAGAAGGCAAACAG
Ptgs2	NM_011198	CCACTTCAAGGGAGTCTGGA
		AGTCATCTGCTACGGGAGGA
Stat1	NM_009283	TCCATCGAGCTCACTCAGAA
		TGTTCCAACTCCTCCAGCTT
Tbp	NM_013684	CTGGAATTGTACCGCAGCTT
		TCCTGTGCACACCATTTTTC
Tyk2	NM_018793	TCTAGCGAGGAGGAGATCCA
		GATGTGCTGTCGGAAGGAAT

Supplementary computational methods

HDAC3 +/- and -/- microarray analysis

CEL files were imported in R. Background correction and RMA normalization were performed using the Affy package (Irizarry et al). Log2-transformed data was then used to calculate fold changes. Statistical significance among triplicates was assessed through two-tailed Welch t-tests. Probeset to gene annotations were retrieved from mogene10stv1.r3cdf R annotation package. Every probeset showing at least a 2-fold change between two conditions with a p-value in the two-tailed Welch t-test equal or lower than 0.05 was considered as significantly differentially expressed.

Gene Set enrichment analysis

Gene set enrichment analysis (GSEA, Subramanian et al.) is a computational approach to test if a defined set of genes shows concordant and significant differences between two biological states.

Genes whose expression was significantly different among HDAC3 -/- (KO) and HDAC3 +/- (wt) were gathered for the LPS treated (4h) condition. These two groups were then split in up- and down- regulated to finally obtain two gene sets. Transcripts were ranked based on their difference in expression between the LPS treated and the untreated conditions in the wt. Then, for each one of the gene sets defined above, an enrichment score reflecting the degree to which it is over-represented at the extremes (top or bottom) of the ranked list was computed. A p-value for the score was then estimated through the generation of an empirical distribution of scores. This is built running GSEA on 1'000 random datasets obtained permuting the labels of the genes in the original dataset. Finally, the significance level is adjusted for multiple hypotheses testing, depending on the number of gene sets tested.

GSEA R implementation was used to perform these analyses.

Motif analysis (transcriptional start sites of differentially expressed genes)

Position-specific weight matrices (PWMs) were collected from the literature (Portales-Casamar et al, Badis et al, Berger et al, Bucher et al, Jolma et al), and used to build a custom set of 491 models.

For each class of interest, probesets were mapped to RefSeq genes and genomic DNA sequences spanning from -500 to +250 bp from their transcriptional start sites (TSSs) were retrieved. Different comparisons were performed, using in turn each dataset as foreground or as background depending on the type of enrichment we were pursuing. In

each comparison we used Pscan (Zambelli et al.) to detect statistically significant over-represented PWMs. In case a PWM showed a p-value equal or lower than 0.01 (two-tailed Welch's t-test) it was considered as significantly over-represented. The Pscan source code was modified in order to replace the statistical evaluation step based on the z-test with a step based on the t-test. The t-test is more suitable than the z-test when comparing datasets with similar cardinality (see Zambelli et al., supplementary material).

Analysis of microarrays from the literature

The set of Ifnb-regulated genes was determined by the union of two datasets available in the literature, namely a time-course of Ifnb treatment (Raza et al.) and expression data from IFNAR -/- macrophages stimulated with LPS (Cheng et al.).

(Raza et al.) Data was downloaded from the GEO (GSE20403) as a matrix file (RMA normalized data values). Once imported in R, we considered a probeset as differentially expressed when it reaches 1.5-fold change between two conditions, with a p-value equal or lower than 0.05 in a two-tailed Welch t-test. A probeset was considered as Ifnb-regulated if it was found up-regulated at either 1h, 2h or 4h (compared to untreated macrophages). Probesets were annotated to the corresponding genes using the Affymetrix Datasheet (MoEx-1_0-st-v1.na30.mm9).

(Cheng et al.) Data was downloaded from GEO (GSE27112) in SOFT Format. It was imported in R and then analyzed using Limma (Smyth 2004). A probeset was considered as Ifnb-regulated if down-regulated by at least 2-fold in the IFNAR -/- LPS treated 3h compared to its wild-type counterpart (same LPS treatment in wt macrophages). Probeset to gene annotations were retrieved through the R package biomaRt (Durinck et al.).

For both datasets the information was then collapsed to gene symbols and the two sets were merge together to define the Ifnb-regulated genes.

ChIP-seq analysis

Single-end 36-bp reads were aligned to the mm9 genome using Bowtie (Langmead et al.). In order to keep only the reads mapping to a unique position in the genome, Bowtie was run with the option –m 1. The option –v 2 was also specified to allow for a maximum of 2 mismatches per read.

In order to define the H4ac enriched regions (peaks), each HDAC3 -/- sample (untreated or LPS treated, 4h) was compared to its wild-type counterparts using Model-based Analysis for Chip-Seq (MACS, Zhang et al.). In order to define H4-acetylated regions that were unaffected by the KO (in untreated and independently in LPS treated), wt samples were compared with previously published bone marrow derived macrophages input DNA (GEO accession: GSM499415). These lists were then purged from regions that were either more or less acetylated in the KO compared to the wt. Regions were considered significantly enriched at a p-value threshold of 10^{-10}.

MACS is also able to generate raw tracks for visualization on the UCSC genome browser (Fujita et al.). Since the sequencing depths of the different samples were very heterogeneous, all the tracks were re-scaled to 10 millions reads. Gene Interval Notator (GIN), a tool included in the CARPET suite (Cesaroni et al.), was then used to annotate all regions over mm9 UCSC known genes extracted from the UCSC genome browser (Fujita et al.). GIN was run with priority set to "gene" and "-20000" as promoter definition. Pie charts summarizing genomic annotations were generated using R.

Hierarchical clustering of ChIP-seq signals

Raw data for PU.1 in untreated macrophages (GEO accession: GSM487450), H3K4me1 in untreated (GEO accession: GSM487453) and H3K4me3 in untreated and LPS treated

(4h) macrophages (GEO accession: GSM470558, GSM470559) were gathered. ChIP-seq signals for these datasets along with the raw data of the H4ac datasets in this study were quantified into hypo- and hyper-acetylated regions in LPS treated samples and then hierarchically clustered.
First of all, the number of overlapping uniquely alignable reads was computed for each region. These numbers considered any PCR-bias that could have affected each samples individually, so each set was cleaned accordingly. These numbers were then linearly normalized to the number of tags of the largest dataset. Since hyper- and hypo-acetylated regions are heterogeneous in length, these numbers were divided by the individual length in kbp of each region. Finally, numbers were forced in the range 0-1 in an antibody-specific manner. That means that all the values coming from ChIP-seqs performed with the same antibody were considered together when rescaled to the 0-1 intervals. All computations were performed using custom C++ scripts. Regions were then hierarchically clustered with R (average linkage and Pearson correlation as distance measure). Heatmaps were generated with R as well.

Motif analysis (hypo- and hyper-acetylated regions from the ChIP-seq)
Clover (Frith et al.) was used to perform motif over-representation analysis in regions whose acetylation was significantly affected by HDAC3 KO. The analysis was performed on hyper- and hypo-acetylated sets of regions in untreated as well as LPS treated macrophages. Each set was compared to a background of matching acetylated regions that were unaffected by the HDAC3 KO and were found significant against the input DNA (MACS, p<=1e-10). In this analysis we only considered distal regions (distant more than 2.5 Kbp form any UCSC known gene TSS). The analysis was run with the same PWMs used in the previously described Pscan analysis. A PWM was retained only when significantly over-represented (p<=0.01) compared to the matched background. Clover is available as a standalone tool while results were parsed using a custom Python script.

Hyperacetylation (gene bodies) analysis
We retrieved all the genomic regions corresponding to gene bodies of the RefSeq genes in the mm9 genome. For each one of them, the number of raw reads in the region was calculated for each H4ac sample. Two parallel analyses were performed for the untreated and the LPS treated samples. For each region, the number of reads calculated above and the number of reads falling outside the region for each sample were used to calculate the probability of this event to happen by chance using a chi-square test. In this way we were able to assess whether the significance of the difference in acetylation between the KO and the wt, taking into account the distinctive sequencing depth of the two samples. Chi-square p-values were then corrected for multiple hypotheses testing using the Benjamini-Hochberg correction (Benjamini & Hochberg). All gene bodies showing corrected p-values equal or lower than 0.01 and a ratio between the normalized numbers of tags in the KO versus the wt higher than one were considered as hyper-acetylated. This information was then collapsed to gene symbols. A gene symbol was considered hyper-acetylated every time at least one overlapping RefSeq gene body was found significant.

References

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003 Feb 15;31(4):e15.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50. Epub 2005 Sep 30.

Portales-Casamar E et al. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. 2010 Jan;38(Database issue):D105-10. Epub 2009 Nov 11.

Badis G et al. Diversity and complexity in DNA recognition by transcription factors. Science. 2009 Jun 26;324(5935):1720-3. Epub 2009 May 14.

Berger MF et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell. 2008 Jun 27;133(7):1266-76.

Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. Journal of Molecular Biology Volume 212, Issue 4, 20 April 1990, Pages 563-578.

Jolma A et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. 2010 Jun;20(6):861-73.

Zambelli F et al. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. Nucleic Acids Res. 2009 Jul 1;37(Web Server issue):W247-52.

Raza S, McDerment N, Lacaze PA, Robertson K, Watterson S, Chen Y, Chisholm M, Eleftheriadis G, Monk S, O'Sullivan M, Turnbull A, Roy D, Theocharidis A, Ghazal P, Freeman TC. Construction of a large scale integrated map of macrophage pathogen recognition and effector systems. BMC Syst Biol. 2010 May 14;4:63.

Christine S. Cheng, Kristyn E. Feldman, James Lee, Shilpi Verma, De-Bin Huang, Kim Huynh, Mikyoung Chang, Julia V. Ponomarenko, Shao-Cong Sun, Chris A. Benedict, Gourisankar Ghosh, and Alexander Hoffmann. The Specificity of Innate Immune Responses Is Enforced by Repression of Interferon Response Elements by NF-kB p50. Sci. Signal., 22 February 2011 Vol. 4, Issue 161, p. ra11

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3, No. 1, Article 3.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005 Aug 15;21(16):3439-40.

Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res. 2004 Feb 26;32(4):1372-81. Print 2004.

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009, 10:R25.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008, 9:R137.

Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. The UCSC Genome Browser database: update 2011. Nucleic Acids Res 2010, 39:D876-882.

Cesaroni M, Cittaro D, Brozzi A, Pelicci PG, Luzi L.CARPET: a web-based package for the analysis of ChIP-chip and expression tiling data. Bioinformatics 2008, 24:2918-2920.

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach for Multiple Testing. 1995